Exhibit A

The leucine-rich repeat as a protein recognition motif

Bostjan Kobe* and Andrey V Kajava†

Leucine-rich repeats (LRRs) are 20–29-residue sequence motifs present in a number of proteins with diverse functions. The primary function of these motifs appears to be to provide a versatile structural framework for the formation of protein–protein interactions. The past two years have seen an explosion of new structural information on proteins with LRRs. The new structures represent different LRR subfamilies and proteins with diverse functions, including GTPase-activating protein rna1p from the ribonuclease-inhibitor-like subfamily; spliceosomal protein U2A′, Rab geranylgeranyltransferase, internalin B, dynein light chain 1 and nuclear export protein TAP from the SDS22-like subfamily; Skp2 from the cysteine-containing subfamily; and YopM from the bacterial subfamily. The new structural information has increased our understanding of the structural determinants of LRR proteins and our ability to model such proteins with unknown structures, and has shed new light on how these proteins participate in protein–protein interactions.

Addresses
*Department of Biochemistry and Molecular Biology/Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia; e-mail: kobe@biosci.uq.edu.au
†Center for Molecular Modeling, Center for Information Technology, National Institutes of Health, Bethesda, MD 20892, USA; e-mail: kajava@helix.nih.gov

Current Opinion in Structural Biology 2001, **11**:725–732

0959-440X/01/$ – see front matter

Abbreviations

CC	cysteine-containing
CTE	constitutive transport element
InlB	internalin B
LC1	light chain 1
LRR	leucine-rich repeat
LRV	leucine-rich repeat variant
PDB	Protein Data Bank
RabGGT	Rab geranylgeranyltransferase
RI	ribonuclease inhibitor
RNP	ribonucleoprotein
snRNA	small nuclear RNA
TpLRR	*Treponema pallidum* LRR

Introduction

Repeating amino acid segments are increasingly recognized as important components of proteins, particularly eukaryotic ones [1,2]. A subset of repeating motifs correspond to structural units that assemble in a superhelical fashion and form solenoid protein structures [3,4•,5•].

One such repeating motif was first recognized in the leucine-rich α2-glycoprotein and was termed the leucine-rich repeat (LRR) [6]. An ever-increasing number of proteins with diverse functions have subsequently revealed tandem arrays of related amino acid motifs (reviewed in [3,7–10]). Most but not all of these proteins are eukaryotic and most if not all appear to be involved in protein–protein recognition processes. The LRRs are generally 20–29 residues long and contain a conserved 11-residue segment with the consensus sequence $LxxLxLxx^{N}/_{C}xL$ (x can be any amino acid and L positions can also be occupied by valine, isoleucine and phenylalanine). The crystal structure of ribonuclease inhibitor (RI) yielded the first insight into the 3D structural arrangement of LRRs [11] and, soon thereafter, crystal structures of complexes of RI with its ligands provided the first structural views revealing how the LRR structure is used as a protein recognition motif [12,13]. This structural information also formed the basis of numerous attempts to model LRR proteins and their ligand complexes [10,14–23].

In the past few years, a number of new 3D structures of LRR proteins have been determined. In this review, we focus on this novel structural information and its implications for understanding the structural and functional attributes of LRR proteins.

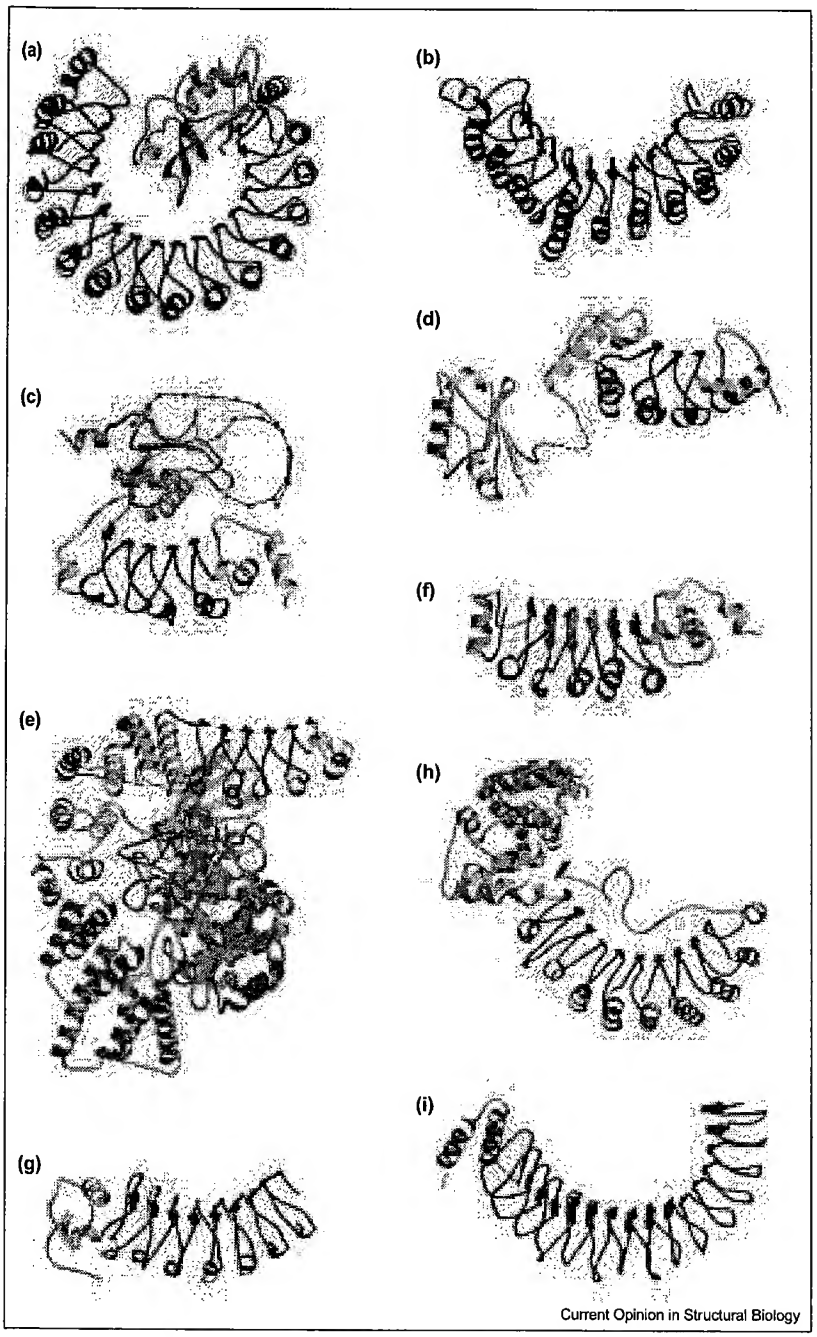
New structures of leucine-rich repeat proteins

The structure of porcine RI showed that LRRs corresponded to structural units, each consisting of a β strand and an α helix connected by loops [11]. The structural units were arranged so that all the strands and helices were parallel to a common axis, resulting in a nonglobular, horseshoe-shaped molecule with a curved parallel β sheet lining the inner circumference of the horseshoe and the helices flanking the outer circumference (Figure 1a).

The structure of RI explained the conservation of residues that constitute an LRR. The conserved pattern $LxxLxLxx^{N}/_{C}xL$ corresponded to the segment surrounding the β strands. The available structure and sequence information strongly suggested that other proteins containing LRRs could have structures related to RI, but that substantial structural variability may exist in the regions between the β strands ('interstrand' regions). It was proposed that the helical region may be shorter or even substituted with an extended structure in some cases [7,10]. The latter possibility even led to speculation that shorter LRRs (all LRRs in RI are 28–29 residues long) may have structures that are more similar to the β helix of pectate lyase [24] than to the β/α horseshoe of RI [8,9,25,26] (see also Update).

Recently, the 3D structures of several new LRR proteins have been determined [27,28••–34••] (Figure 1, Table 1). The new structures are particularly informative because of the diversity of lengths and sequences of the individual LRRs in these proteins. The structures all show significant similarities, including a curved overall shape with a parallel β sheet on the concave side and mostly helical elements on the convex side.

Figure 1



3D structures of LRR proteins. The LRR domains are shown in cyan, the flanking regions that are an integral part of the LRR domain but do not correspond to LRR motifs are shown in gray, and the other domains/subunits in the structure are shown in magenta. **(a)** RI (PDB code 2BNH), **(b)** rna1p (PDB code 1YRG), **(c)** U2A′–U2B″ (PDB code 1A9N), **(d)** TAP (PDB code 1FO1), **(e)** RabGGT (PDB code 1DCE), **(f)** dynein LC1 (PDB code 1DS9), **(g)** InlB (PDB code 1D0B), **(h)** Skp2–Skp1 (PDB code 1FQV) and **(i)** YopM (PDB code 1G9U).

The structure most similar to RI is that of the GTPase-activating protein rnalp [28••] (Figure 1b). This protein stimulates the GTPase activity of the protein Ran, which is involved in nucleocytoplasmic transport processes. Its LRRs are similar to those of RI, but much more irregular. LRRs 1, 3 and 5 differ significantly from the typical LRR pattern; the helix in LRR5 is 22 residues long (10–14 residues in most rna1p LRRs), LRR1 contains an insertion in the β–α loop that results in the formation of an additional small β sheet and the shortening of the usual α helix, and the β–α loop of LRR3 has a six-residue insertion that protrudes from the side of the horseshoe. Mutagenesis studies suggest that an arginine residue in this insertion may be involved in binding Ran and stimulating its GTPase activity [28••].

The structure of the ternary complex between fragments of the spliceosomal proteins U2B″ (comprising a ribonucleoprotein [RNP] domain) and U2A′ (containing LRRs), and

Table 1

3D structures of LRR proteins.

LRR protein	Organism	Ligand present in structure	Function	Number of LRRs	LRR length (residues)	LRR subfamily	Secondary structure in interstrand segment	PDB code	Refere -nces
RI	Pig	–	Ribonuclease inhibitor	15	28–29	RI-like	α helix	2BNH	[11]
RI	Pig	Ribonuclease A	Ribonuclease inhibitor	15	28–29	RI-like	α helix	1DFJ	[12]
RI	Human	Angiogenin	Ribonuclease inhibitor	15	28–29	RI-like	α helix	1A4Y	[13]
rna1p	*S. pombe*	–	GTPase-activating protein for Ran	11	28–37	RI-like	α helix	1YRG	[28••]
U2A′	Human	U2B′ snRNA	Splicing	5	22–26	SDS22-like	3_{10} helix, α helix, extended	1A9N	[27]
TAP	Human	–	RNA export from nucleus	4	24–41	SDS22-like	α helix	1FO1	[33••]
RabGGT	Rat	–	Rab geranylgeranyl transferase	5	22–27	SDS22-like	3_{10} helix, α helix	1DCE	[29••]
LC1 of dynein	*C. reinhardtii*	–	Protein–protein interactions in molecular motor complex	6	22–25	SDS22-like	α helix	1DS9	[31••]
InlB	*L. monocyto-genes*	–	Phagocytosis	7.5	22	SDS22-like	3_{10} helix	1D0B	[30••]
Skp2	Human	Skp1	Substrate binding in ubiquitination	10	23–27	Cysteine-containing	α helix	1FQV	[32••]
YopM	*Y. pestis*	–	Virulence factor	15	20–22	Bacterial	Polyproline II	1G9U	[34••]

a hairpin loop of U2 small nuclear RNA (snRNA) provided the first view of shorter and highly imperfect LRRs [27] (Figure 1c). The interstrand segments form either 3_{10}-helical, α-helical or more irregular extended structures. Although U2B″ contains the primary RNA-binding interface, U2A′ is absolutely required for cognate RNA binding. U2A′ has a large interface with U2B″, but also interacts directly with the double-stranded stem of U2 snRNA through its basic C-terminal region.

A combination of RNP and LRR domains were also revealed by the structure of a fragment of the TAP protein, suggesting functional similarities with the U2B″–U2A′ system [33••] (Figure 1d). TAP is implicated in mRNA export from the nucleus and is specifically used by simian type D retroviruses to export their unspliced genomic RNA into the cytoplasm of the host cell; TAP directly recognizes the constitutive transport element (CTE) of retroviral RNAs. The crystal structure of the minimal CTE-binding fragment of TAP reveals that the RNP and LRR domains are very loosely associated; in fact, unambiguous identification of the domains from a single molecule has not been possible because of the disorder of the interdomain linker and the nature of the crystal packing.

Rab geranylgeranyltransferase (RabGGT) catalyses the addition of two geranylgeranyl groups to the C terminus of Rab proteins, resulting in the membrane association that is required for the function of these proteins in intracellular vesicular trafficking. The crystal structure reveals three distinct structural modules in the α subunit and one compact domain in the β subunit [29••] (Figure 1e). The LRR domain is part of the α subunit and appears to be rigidly attached to the helical domain of the same subunit. The interstrand segment contains an α helix in the last repeat and 3_{10} helices in all the other LRRs.

The structure of light chain 1 (LC1) of the molecular motor dynein from the alga *Chlamydomonas* has been determined by NMR [31••] (Figure 1f). This protein contains six central LRRs flanked by helical domains. Its highest structural similarity is with U2A′ and includes both the LRR domain and the C-terminal helical domain, although the helical domains are oriented differently in the two proteins.

The protein internalin B (InlB) from *Listeria monocytogenes* plays a role in phagocytosis induced by this pathogen in mammalian cells; it specifically activates phosphoinositide-3-kinase by stimulating tyrosine phosphorylation of adaptor proteins. The structure of an N-terminal fragment of InlB that is responsible for these activities shows an LRR domain capped at the N terminus by a calcium-binding region [30••] (Figure 1g).

F-box proteins are characterized by an ~40-residue F-box motif linked to a protein–protein interaction module such as the LRR or the WD-40 repeat domain. The F-box protein Skp2 contains an LRR domain and regulates the G1/S transition in mammalian cells by controlling the degradation, via ubiquitination, of the cyclin-dependent protein kinase inhibitor $p27^{Kip1}$. A crystal structure of the complex between the ubiquitin-protein ligase components Skp1 and

Table 2

Subfamilies of LRR proteins.

LRR subfamily	Typical LRR length (range)	Organism origin	Cellular location	Structures available
RI-like	28–29 (28–29)	Animals	Intracellular	RI, rna1p
SDS22-like	22 (21–23)	Animals, fungi	Intracellular	U2A', TAP, RabGGT, LC1, InlB
Cysteine-containing	26 (25–27)	Animals, plants, fungi	Intracellular	Skp2
Bacterial	20 (20–22)	Gram-negative bacteria	Extracellular	YopM
Typical	24 (20–27)	Animals, fungi	Extracellular	No
Plant-specific	24 (23–25)	Plants, primary eukaryotes	Extracellular	No
TpLRR	23 (23–25)	Bacteria	Extracellular	No

Consensus sequence*

RI-like	x x x **L** x x **L** x **L** x x **N/C** x **L** x x x g **o** x x **L** x x **o** **L** x– x
SDS22-like	**L** x x **L** x x **L** x **L** x x **N** x **I** x x **I** x x **L** x– x
Cysteine-containing	**c** x x **L** x x **L** x **L** x x **c** x– x **I** T D x x **o** x x **L** **a** x– x
Bacterial	P x x **L** x x **L** x **V** x x **N** x **L** x x **L** P e/d **L** –
Typical	**L** x x **L** x x **L** x **L** x x **N** x **L** x x **L** p x x **o** **F** x– x
Plant-specific	**L** x– x **L** x x **L** x **L** x x **N** x **L** t/s g– x **I** P x x **L** **G** x
TpLRR	**C/N** x– x **L** x x **I** x **L** x– x x **L** x x **I** g x x **A** **F** x x

*Residues identical or conservatively substituted in more than 50% and 30% of the repeats of a given protein are shown in uppercase and lowercase, respectively. Residues directed into the interior of the known protein structures or models are shown in boxes in bold. '–', possible insertion site; o, a nonpolar residue; x, any residue.

Skp2 has been described [32••] (Figure 1h). The LRR domain adopts ten LRRs and a C-terminal tail folds back towards the first LRR.

The plague virulence protein YopM from *Yersinia pestis* contains short 20-residue LRRs. In YopM, the variable interstrand regions adopt a polyproline II helical conformation interrupted by one or two residues in α-helical conformation [34••] (Figure 1i). The N-terminal edge of the LRR domain is flanked by an α-helical hairpin. Individual YopM molecules wrap around each other and form tetramers in the crystal.

Most known LRR solenoids show very little twist. However, some degree of twisting is observed in InlB and particularly YopM; the reasons are not yet clear. It is, however, now clear that all the major classes of LRR have curved horseshoe structures closely resembling the RI solenoid, not the β helix of pectate lyase. The curvature of the LRR structures is defined, in part, by the conformation of the interstrand segments, which defines the thickness of the wedge-shaped units containing a β strand on one side and a helix or more extended structure on the other. Molecular modeling further shows that curvature is also imposed by the bulky nonpolar residue and asparagine flanking the β strand (italic in the *L*xxLxLxx*N*xL consensus pattern), which are directed into the interior of the structure [10]. Although the structural units of β-helical proteins have a sequence motif very similar to an LRR, they lack this bulky nonpolar residue and instead typically contain a small sidechain in the equivalent position; consequently, β-helix structures are not curved.

The structures show that the more irregular the LRRs, the more irregular their 3D conformations. Similarly, leucines are highly preferred in the rather regular LRRs, such as in RI, but other hydrophobic residues, such as isoleucine, valine and phenylalanine, substitute more frequently in the more irregular repeats.

The high-resolution structures of InlB and human RI suggest that water molecules serve as important structural elements in LRR proteins, by forming bridging hydrogen bonds between adjacent repeats. The water molecules are organized in distinct spines along the convex face of the LRR structure.

Subfamilies of leucine-rich repeat proteins

Sequence analyses of LRR proteins suggested the existence of several different subfamilies of LRRs [9,10,14,26,35]. The most recent classifications reveal at least seven distinct subfamilies (Table 2; [10]; AV Kajava, B Kobe, unpublished data). The significance of this classification is that repeats from different subfamilies never occur simultaneously in the same protein and have most probably evolved independently. The known structure of RI allowed the construction of 3D models of LRRs from the other subfamilies ([10,14]; AV Kajava, B Kobe, unpublished data). The new crystal and NMR structures now provide experimental structural information on three of the remaining subfamilies.

RI and rna1p belong to the RI-like subfamily. This is a minor subfamily characterized by longer, typically 28- or 29-residue repeats. The major attributes are the presence of an α helix in the variable part of the LRR, large curvature compared to other LRR proteins and essentially no twist of the parallel β sheet.

Sequence analyses suggest that U2A′, TAP, RabGGT, dynein LC1 and InlB all belong to the LRR subfamily represented by the protein SDS22+ [10]. Inspection of these structures confirms that the less perfect the repetition of the sequence motif, the less regular the structure. In three dimensions, neither U2A′, RabGGT nor dynein LC1 have two identical LRR conformations within one domain. By contrast, InlB has seven well-conserved 22-residue repeats and the structure contains several repeats with the same conformation. The backbones of the 22-residue repeats from U2A′, RabGGT and InlB are also very similar. Surprisingly, however, some repeats that fit the consensus sequence of the SDS22-like LRRs show considerable differences in the backbone conformations of their interstrand regions, depending on the surrounding protein context.

The crystal structure of the Skp2 LRR domain [32••] provides the first view of a protein from the cysteine-containing (CC) LRR subfamily [10]. LRRs from Skp2 vary in length between 23 and 27 residues, but, despite these length differences, the convex part of the horseshoe structure is formed by similar α helices containing 2–3 turns. The α helix of the CC LRR is shifted relative to the LRR of RI.

Finally, the crystal structure of YopM from *Y. pestis* shows, for the first time, the architecture of a bacterial LRR [34••]. The variable region in the LRRs of this protein adopts a more extended conformation, similar to the polyproline II helix. Experimental structural information is still lacking for the plant-specific, as well as the most populated typical, LRR subfamily.

The new structural information on LRR proteins and molecular modeling suggest that the characteristic horseshoe structure may not require the presence of the characteristic conserved asparagine/cysteine. A modified sequence profile search (AV Kajava, B Kobe, unpublished data) revealed a group of bacterial proteins exemplified by a protein from *Treponema pallidum* (TpLRR) containing 23-residue repeats with the consensus pattern LxxLxLxxxLxxIgxxAFxx$^{C}/_{N}$xx (Table 2). Modeling suggests that proteins with such 'inverted' $^{C}/_{N}$xxLxxLxL motifs can adopt a similar horseshoe structure to the more typical LRR proteins. Sequence profiles of known LRR subfamilies are available on the World Wide Web at http://cmm.info.nih.gov/kajava/lrrprofiles.

A protein from *Azotobacter vindelandii* contains a repeating pattern very similar to LRRs. However, the structure of this leucine-rich repeat variant (LRV) protein clearly demonstrates that it is structurally distinct from other LRR proteins [3,36]. In LRV, the usual LRR consensus sequence (LxxLxLxx$^{N}/_{C}$xL) is replaced by ExLxxLxxDxD; most importantly, a leucine residue crucial to the formation of the β strand relocates by one residue. This transition transforms the usual β strand into a 3_{10} helix, resulting in a remarkably different solenoid fold containing structural units composed of two antiparallel helices.

Flanking regions

In a regular LRR structure, the hydrophobic core would be exposed to solvent at the ends. Most LRR proteins therefore contain flanking regions that are an integral part of the LRR domain.

In the structure of U2A′, an amphipathic α helix shields the hydrophobic core of the C-terminal LRR, followed by a loop that aligns with the β strand of the LRR [27]. A sequence motif (Yrxxoxxxo-Pxo-xxLD; 'o' is a hydrophobic residue, '-' is a possible insertion site) corresponding to this C-terminal flanking motif has been recognized in several LRR proteins and is termed the 'LRR cap' [37]. Indeed, the structures of these motifs in U2A′, TAP, RabGGT and dynein LC1 are remarkably similar.

InlB has a hydrophilic 'N-terminal cap' that protects the first LRR and functions to bind calcium ions. In addition to the capping role, this region may have a functional role in cell invasion.

In extracellular proteins, LRRs are often flanked by cysteine-rich domains. Sequence analyses revealed four characteristic C-terminal (C-flanking) domains and one N-terminal (N-flanking) domain [10]. The most common C-flanking domain (CF1) contains four cysteines, a small proteoglycan-specific domain (CF2) contains two cysteines, a G protein coupled receptor specific domain (CF3) contains three cysteines and a plant LRR protein specific domain (CF4) contains two cysteines. The N-flanking domain (NF) contains four cysteines and is found in proteins with typical LRRs that may have CF1, CF2 or CF3 domains on the C-terminal side. No structural information is currently available for any of these cysteine-rich domains.

Protein–protein interactions

A survey of the functions of LRR proteins suggests that the major function of the LRRs may be to provide a structural framework for the formation of protein–protein interactions. However, direct structural information on how LRR proteins recognize their binding partners is only available for two proteins, RI and U2A′.

Ribonuclease A [12] and angiogenin [13] bind to a very similar site on the RI molecule, involving mainly residues that lie on β strands and β–α loops. Although the two ligands are distantly related and have overlapping binding sites, mutagenesis studies suggest that different specific contacts are in fact important for the binding of the two proteins [38].

U2B″ binds to U2A′ via a helix from the RNP domain that lies on the concave surface of the parallel β sheet of U2A′ LRRs. Further contacts come from outside the LRR region. Interestingly, the protein U1A, which contains an RNP domain closely related to that of U2B″, does not bind U2A′, but can form a stable complex after two subtle mutations (D24E and L28R) are introduced [39]. The intriguing similarity between the U2A′–U2B″ system and the TAP protein suggests similarities in RNA binding [33••]. Mutagenesis confirmed that the N-terminal domain of TAP functions as a *bona fide* RNP domain and that the LRR domain does not show general RNA-binding activity, but is essential for CTE binding. However, the roles of the LRR domains in U2A′ and TAP are at least partly divergent; the recognition of the CTE by TAP requires the RNP and LRR domains to be part of one molecule, whereas U2B″ and U2A′ function as separate proteins. However, the LRR domain may more generally act independently to interact with other RNP domain proteins, analogous to the U2A′–U2B″ system.

The approximately 30-residue C-terminal tail of Skp2 was found to extend back towards the first LRR, packing loosely in the concave surface of the LRR domain [32••]. This tail may therefore be involved in or may regulate substrate recognition, or itself be a substrate, and thus represents another example of LRR-mediated interaction.

Possible ligand-binding sites have been mapped onto the surfaces of other LRR proteins with known structures, based on the results of mutagenesis and the conservation of surface-exposed residues in the protein families. In both rna1p and InlB, the β–α loop and the concave surface are the most likely binding regions. LC1 is known to associate with two different proteins using distinct interfaces; a hydrophobic patch on the β-sheet face of the protein is proposed to bind the γ heavy chain, but, unusually, the binding site for the axonemal protein p45 is likely to involve a charged surface on the opposite face.

In summary, the concave face and the adjacent loops are the most common protein interaction surfaces on LRR proteins. The structure of U2B″–U2A′ shows that the concave surface of the LRR domain is ideal for interaction with an α helix and this may be a recurring feature of protein–protein interactions in LRR proteins. The available data continue to support earlier conclusions [8] that the elongated and curved LRR structure, coupled with the observed structural flexibility [12], provides an outstanding framework for achieving diverse protein–protein interactions.

Modeling of leucine-rich repeat structures

The available structural information can be used to construct models of novel LRR proteins. No experimental information is available yet for the structures of the typical, plant-specific and TpLRR subclasses. However, analyses suggest that quite reliable models for all the major classes can be obtained using the available information ([10]; AV Kajava, B Kobe, unpublished data). In particular, a comparison of models constructed solely on the basis of the structure of RI [10,14] with the new experimental structures uncovers the strengths and weaknesses of LRR modeling, which should be considered when these models are used (AV Kajava, B Kobe, unpublished data). The comparison suggests that the general architecture, curvature, 'interior/exterior' orientations of sidechains and even backbone conformations of the LRR structures can be predicted correctly. What remain difficult to predict correctly are the twist of the overall solenoid structure and the sidechain rotamers. The reliability of LRR protein modeling suggests that it would be informative to apply similar modeling approaches to other classes of solenoid proteins.

The mapping of residues that are conserved within a protein family onto the surface of the LRR protein structure can shed light on the regions possibly involved in protein–protein interactions. One has to be careful, however, to exclude the surface-exposed residues that may be conserved for structural and other reasons.

Functions of leucine-rich repeat proteins

LRR proteins participate in many biologically important processes, such as hormone–receptor interactions, enzyme inhibition, cell adhesion and cellular trafficking. A number of recent studies revealed the involvement of LRR proteins in early mammalian development [40], neural development [41], cell polarization [42], regulation of gene expression [43] and apoptosis signaling [44]. It was shown that LRR domains may be critical to the morphology and dynamics of the cytoskeleton [31••,45]. In all these processes, the LRR domains probably mediate protein–protein interactions. The exception is a carrot LRR protein that may use its repetitive structure to inhibit ice crystallization [46]. Apart from the LRRs providing an ideal structural framework for achieving protein–protein interactions, the repetitive structure may be beneficial in processes in which the rapid generation of new variants is required, such as in plant disease resistance [35] or bacterial virulence [30••,34••], because it can evolve more rapidly [2,4•].

Conclusions

The identification of new LRR proteins through genome sequencing projects and the functional characterization of new and old LRR proteins emphasize the important roles LRR domains play in various cellular processes and suggest that the major function of LRR domains is to facilitate protein–protein interactions. The new structural information strongly suggests that LRR proteins from all the major subfamilies have related structures and facilitates reliable modeling of such proteins with unknown structures. Analyses of the determined or modeled structures in the light of mutational data and/or sequence conservation within a protein family can shed light on the regions involved in the formation of protein–protein interactions.

The major area for which new information is required urgently is structural analysis of LRR protein–ligand complexes. Currently, 3D structural information on complexes is available for only two systems, RI–ribonuclease and U2A′–U2B″. Although this allows some general conclusions to be drawn regarding the structural basis of protein–protein interactions involving LRR proteins, experimental data are required to test the hypotheses and help us understand the functions of this important group of proteins.

Update

Ward and Garrett [47] recently suggested, based on sequence and structure comparisons, that pectate lyase and the L domains of members of the insulin receptor and epidermal growth factor receptor families are members of the LRR superfamily. We discussed above how these structures and the corresponding repeat profiles differ from the LRR proteins.

The structure of a larger fragment of InlB and an equivalent fragment of the related InlH have recently been reported [48]. The structures show that the so-called 'inter-repeat region' C-terminal to the LRR region corresponds to an immunoglobulin-like domain contiguously fused to the LRR domain.

Acknowledgements

We thank David S Waugh and co-workers for sharing unpublished data on the YopM protein. BK is a Wellcome Senior Research Fellow in Medical Science in Australia.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Heringa J: **Detection of internal repeats: how common are they?** *Curr Opin Struct Biol* 1998, **8**:338-345.

2. Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D: **A census of protein repeats.** *J Mol Biol* 1999, **293**:151-160.

3. Kobe B: **Leucines on a roll.** *Nat Struct Biol* 1996, **3**:977-980.

4. • Kobe B, Kajava AV: **When protein folding is simplified to protein coiling: the continuum of solenoid protein structures.** *Trends Biochem Sci* 2000, **25**:509-515.
This review discusses the basic structural features of solenoid proteins, similarities and differences between different solenoids, what distinguishes them from other protein structures and how this relates to their functional properties.

5. • Kajava AV: **Proteins with repeated sequence: structural prediction and modeling.** *J Struct Biol* 2001, **134**:132-144.
This paper discusses the relationship between the amino acid sequence and 3D structure of proteins with internal repeats. The analyses suggest rules that can be used for structure prediction of repeat-containing proteins.

6. Takahashi N, Takahashi Y, Putnam FW: **Periodicity of leucine and tandem repetition of a 24-amino acid segment in the primary structure of leucine-rich α2-glycoprotein of human serum.** *Proc Natl Acad Sci USA* 1985, **82**:1906-1910.

7. Kobe B, Deisenhofer J: **The leucine-rich repeat: a versatile binding motif.** *Trends Biochem Sci* 1994, **19**:415-421.

8. Kobe B, Deisenhofer J: **Proteins with leucine-rich repeats.** *Curr Opin Struct Biol* 1995, **5**:409-416.

9. Buchanan SG, Gay NJ: **Structural and functional diversity in the leucine-rich repeat family of proteins.** *Prog Biophys Mol Biol* 1996, **65**:1-44.

10. Kajava AV: **Structural diversity of leucine-rich repeat proteins.** *J Mol Biol* 1998, **277**:519-527.

11. Kobe B, Deisenhofer J: **Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats.** *Nature* 1993, **366**:751-756.

12. Kobe B, Deisenhofer J: **A structural basis of the interactions between leucine-rich repeats and protein ligands.** *Nature* 1995, **374**:183-186.

13. Papageorgiou AC, Shapiro R, Acharya KR: **Molecular recognition of human angiogenin by placental ribonuclease inhibitor — an X-ray crystallographic study at 2.0 Å resolution.** *EMBO J* 1997, **16**:5162-5177.

14. Kajava AV, Vassart G, Wodak SJ: **Modeling of the three-dimensional structure of proteins with the typical leucine-rich repeats.** *Structure* 1995, **3**:867-877.

15. Jiang X, Dreano M, Buckler DR, Cheng S, Ythier A, Wu H, Hendrickson WA, El Tayar N: **Structural predictions for the ligand-binding region of glycoprotein hormone receptors and the nature of hormone-receptor interactions.** *Structure* 1995, **3**:1341-1353.

16. Moyle WR, Campbell RK, Rao SNV, Ayad NG, Bernard MP, Han Y, Wang Y: **Model of human chorionic gonadotropin and lutropin receptor interaction that explains signal transduction of the glycoprotein hormones.** *J Biol Chem* 1995, **270**:20020-20031.

17. Bhowmick N, Huang J, Puett D, Isaacs NW, Lapthorn AJ: **Determination of residues important in hormone binding to the extracellular domain of the luteinizing hormone/chorionic gonadotropin receptor by site-directed mutagenesis and modeling.** *Mol Endocrinol* 1996, **10**:1147-1159.

18. Weber IT, Harrison RW, Iozzo RV: **Model structure of decorin and implications for collagen fibrillogenesis.** *J Biol Chem* 1996, **271**:31767-31770.

19. Iozzo RV: **The family of the small leucine-rich proteoglycans: key regulators of matrix assembly and cellular growth.** *Crit Rev Biochem Mol Biol* 1997, **32**:141-174.

20. Janosi JB, Ramsland PA, Mott MR, Firth SM, Baxter RC, Delhanty PJ: **The acid-labile subunit of the serum insulin-like growth factor-binding protein complexes. Structural determination by molecular modeling and electron microscopy.** *J Biol Chem* 1999, **274**:23328-23332.

21. Montgomery J, Ilg T, Thompson JK, Kobe B, Handman E: **Identification and predicted structure of a leucine-rich repeat motif shared by *Leishmania major* proteophosphoglycan and parasite surface antigen 2.** *Mol Biochem Parasitol* 2000, **107**:289-295.

22. Hobby P, Wyatt MK, Gan W, Bernstein S, Tomarev S, Slingsby C, Wistow G: **Cloning, modeling, and chromosomal localization for a small leucine-rich repeat proteoglycan (SLRP) family member expressed in human eye.** *Mol Vis* 2000, **6**:72-78.

23. Hines J, Skrzypek E, Kajava AV, Straley SC: **Structure-function analysis of *Yersinia pestis* YopM's interaction with alpha-thrombin to rule on its significance in systemic plague and to model YopM's mechanism of binding host proteins.** *Microb Pathog* 2001, **30**:193-209.

24. Yoder MD, Keen NT, Jurnak F: **New domain motif: the structure of pectate lyase C, a secreted plant virulence factor.** *Science* 1993, **260**:1503-1507.

25. Heffron S, Moe GR, Sieber V, Mengaud J, Cossart P, Vitali J, Jurnak F: **Sequence profile of the parallel beta helix in the pectate lyase superfamily.** *J Struct Biol* 1998, **122**:223-235.

26. Claudianos C, Campbell HD: **The novel flightless-I gene brings together two gene families, actin-binding proteins related to gelsolin and leucine-rich-repeat proteins involved in Ras signal transduction.** *Mol Biol Evol* 1995, **12**:405-414.

27. Price SR, Evans PR, Nagai K: **Crystal structure of the spliceosomal U2B″-U2A′ protein complex bound to a fragment of U2 small nuclear RNA.** *Nature* 1998, **394**:645-650.

28. Hillig RC, Renault L, Vetter IR, Drell T IV, Wittinghofer A, Becker J:
•• **The crystal structure of rna1p: a new fold for a GTPase-activating protein.** *Mol Cell* 1999, **3**:781-791.
The crystal structure of the GTPase-activating protein rna1p is described. The protein contains 11 RI-like LRRs. Structure analyses and mutagenesis studies suggest probable Ran GTPase binding and activating regions.

29. Zhang H, Seabra M, Deisenhofer J: **Crystal structure of Rab**
•• **geranylgeranyltransferase at 2.0 Å resolution.** *Structure* 2000, **8**:241-251.
The crystal structure of RabGGT reveals a four-domain protein. Two of these domains, the immunoglobulin and LRR domains, may function in binding to the Rab escort protein, which is essential for catalytic activity.

30. Marino M, Braun L, Cossart P, Ghosh P: **Structure of InlB**
•• **leucine-rich repeats, a domain that triggers host cell invasion by the bacterial pathogen *L. monocytogenes*.** *Mol Cell* 1999, **4**:1063-1072.
The crystal structure of an N-terminal fragment of InlB that is implicated in phagocytosis is described. Unexpectedly, a calcium-binding region is found N-terminal to the LRR domain and suggests a role for calcium in phagocytosis.

31. Wu H, Maciejewski MW, Marintchev A, Benashski SE, Mullen GP,
•• King SM: **Solution structure of a dynein motor domain associated light chain.** *Nat Struct Biol* 2000, **7**:575-579.
The solution structure of LC1 of the molecular motor dynein from the alga *Chlamydomonas* is described. This protein is a part of a multisubunit complex and the binding sites for two known binding partners are suggested.

32. Schulman BA, Carrano AC, Jeffrey PD, Bowen Z, Kinnucan ER,
•• Finnin MS, Elledge SJ, Harper JW, Pagano M, Pavletich NP: **Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex.** *Nature* 2000, **408**:381-386.
The crystal structure of the complex between ubiquitin-protein ligase components Skp1 and Skp2 is described. Skp2 contains an N-terminal F-box and a C-terminal LRR domain that is putatively involved in substrate recruitment. This is the first structure of an LRR protein from the CC subfamily.

33. Liker E, Fernandez E, Izaurralde E, Conti E: **The structure of the**
•• **mRNA export factor TAP reveals a *cis* arrangement of a non-canonical RNP domain and an LRR domain.** *EMBO J* 2000, **19**:5587-5598.
The crystal structure of a fragment of the TAP protein that is responsible for binding the CTE of retroviral RNAs is described. The structure shows two domains, an N-terminal RNP domain that was not recognized from the sequence and a C-terminal LRR domain. The domain structure suggests functional similarities with the U2B″–U2A′ spliceosomal complex; the LRR domain may function as a protein–protein interaction module in the export of mRNA from the nucleus.

34. Evdokimov AG, Anderson DE, Routzahn KM, Waugh DS: **Unusual**
•• **molecular architecture of the *Yersinia pestis* cytotoxin YopM: a leucine-rich repeat protein with the shortest repeating unit.** *J Mol Biol* 2001, **312**:807-821.
The crystal structure of the virulence factor YopM from *Y. pestis* is described. This is the first structure of an LRR protein from the bacterial subfamily.

35. Jones DA, Jones JDG: **The role of leucine-rich repeat proteins in plant defences.** *Adv Bot Res* 1997, **24**:89-167.

36. Peters JW, Stowel MHB, Rees DC: **A leucine-rich repeat variant with a novel repetitive protein structural motif.** *Nat Struct Biol* 1996, **3**:1181-1187.

37. Ceulemans H, De Maeyer M, Stalmans W, Bollen M: **A capping domain for LRR protein interaction modules.** *FEBS Lett* 1999, **456**:349-351.

38. Chen CZ, Shapiro R: **Site-specific mutagenesis reveals differences in the structural bases for tight binding of ribonuclease inhibitor to angiogenin and ribonuclease A.** *Proc Natl Acad Sci USA* 1997, **94**:1761-1766.

39. Scherly D, Dathan NA, Boelens W, van Venrooij WJ, Mattaj IW: **The U2B″ RNP motif as a site of protein–protein interaction.** *EMBO J* 1990, **9**:3675-3681.

40. Tong ZB, Nelson LM, Dean J: **Mater encodes a maternal protein in mice with a leucine-rich repeat domain homologous to porcine ribonuclease inhibitor.** *Mamm Genome* 2000, **11**:281-287.

41. Mutai H, Toyoshima Y, Sun W, Hattori N, Tanaka S, Shiota K: **PAL31, a novel nuclear protein, expressed in the developing brain.** *Biochem Biophys Res Comm* 2000, **274**:427-433.

42. Bilder D, Perrimon N: **Localization of apical epithelial determinants by the basolateral PDZ protein Scribble.** *Nature* 2000, **403**:676-680.

43. Linhoff MW, Harton JA, Cressman DE, Martin BK, Ting JP: **Two distinct domains within CIITA mediate self-association: involvement of the GTP-binding and leucine-rich repeat domains.** *Mol Cell Biol* 2001, **21**:3001-3011.

44. Inohara N, Koseki T, del Peso L, Hu Y, Yee C, Chen S, Carrio R, Merino J, Liu D, Ni J *et al.*: **Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor-kappaB.** *J Biol Chem* 1999, **274**:14560-14567.

45. Xu P, Mitchelhill KI, Kobe B, Kemp BE, Zot HG: **The myosin-I-binding protein Acan125 binds the SH3 domain and belongs to the superfamily of leucine-rich repeat proteins.** *Proc Natl Acad Sci USA* 1997, **94**:3685-3690.

46. Worrall D, Elias L, Ashford D, Smallwood M, Sidebottom C, Lillford P, Telford J, Holt C, Bowles D: **A carrot leucine-rich-repeat protein that inhibits ice recrystallization.** *Science* 1998, **282**:115-117.

47. Ward CW, Garrett TP: **The relationship between the L1 and L2 domains of the insulin and epidermal growth factor receptors and leucine-rich repeat modules.** *BMC Bioinformatics* 2001, **2**:4.

48. Schubert WD, Gobel G, Diepholz M, Darji A, Kloer D, Hain T, Chakraborty T, Wehland J, Domann E, Heinz DW: **Internalins from the human pathogen *Listeria monocytogenes* combine three distinct folds into a contiguous internalin domain.** *J Mol Biol* 2001, **312**:783-794.